Comparative Mapping and Targeted-Capture Sequencing of the Gametocidal loci in *Aegilops sharonensis*

Surbhi Grewal,* Laura-Jayne Gardiner, Barbora Ndreca, Emilie Knight, Graham Moore, Ian P. King and Julie King

S. Grewal, B. Ndreca, I. P. King and J. King, Division of Plant and Crop Sciences, School of Biosciences, The University of Nottingham, Sutton Bonington Campus, Loughborough, LE12 5RD, UK; L-J. Gardiner, Institute of Integrative Biology, University of Liverpool, Crown Street, Liverpool, L69 7ZB, UK; E. Knight and G. Moore, Crop Genetics Department, John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, UK. Received _____. *Corresponding author (surbhi.grewal@nottingham.ac.uk).

Core Ideas

- Generation of 54 SNP markers in an *Ae. sharonensis* 4S^{shL} introgression segment in wheat
- Identification of 18 candidate genes in *Ae. sharonensis* for the Gametocidal 'breaker' element.
- Production of a valuable genomic resource that will aide in further mapping the Gc locus.

List of Abbreviations

bp: base pair

EMS: Ethyl Methanesulfonate

FISH: Flourescence in situ hybridization

Gc: Gametocidal

GISH: Genomic in situ hybridization

IL: Introgression line

InDel: Insertion/Deletion

PCR: Polymerase Chain Reaction

SNP: Single Nucleotide Polymorphism

Abstract

Gametocidal (Gc) chromosomes or elements in species such as *Aegilops sharonensis* are preferentially transmitted to the next generation through both the male and female gametes when introduced into wheat. Furthermore, any genes, e.g. genes that control agronomically important traits, showing complete linkage with gametocidal elements, are also transmitted preferentially to the next generation without the need for selection. The mechanism for the preferential transmission of the gametocidal elements appears to occur by the induction of extensive chromosome damage in any gametes that lack the gametocidal chromosome in question. Previous studies on the mechanism of the gametocidal action in *Ae. sharonensis* indicates that at least two-linked elements are involved. The first, the 'breaker' element, induces chromosome breakage in gametes, which have lost the gametocidal elements while the second, the 'inhibitor' element, prevents the chromosome breakage action of the 'breaker' element in gametes, which carry the Gc elements. In this study, we have used comparative genomic studies to map 54 single nucleotide polymorphism (SNP) markers in an *Ae. sharonensis* $4S^{shL}$ introgression segment in wheat and have also identified 18 candidate genes in *Ae. sharonensis* for the 'breaker' element through targeted sequencing of this $4S^{shL}$ introgression segment. This valuable genomic resource will aide in further mapping the Gc locus that could be exploited in wheat breeding to produce new, superior varieties of wheat.

Introduction

Certain *Aegilops* species have an evolutionary unique mechanism that ensures the transmission of specific regions of their genomes to the next generation when introduced to wheat. This phenomenon was first recognized when specific chromosomes, "Gametocidal" chromosomes ('Gc', (Endo, 1978)) or "Cuckoo" chromosomes (Finch et al., 1984), were observed to be preferentially transmitted through both the male and female gametes in a wheat background.

Gc chromosomes or elements have been identified in several S genome species on different chromosomes (Endo, 1982; Endo, 1985; Maan, 1975; Miller et al., 1982; Tsujimoto, 1994; Tsujimoto and Tsunewaki, 1984) as well as in the C genome (Endo, 1979; Endo, 1988; Endo and Tsunewaki, 1975; Endo and Y., 1978). A gametocidal element has also been reported on chromosome $7el_2$ of *Thinopyrum ponticum* (formerly *Agropyron elongatum*; (Kibirige-Sebunya and Knott, 1983)) and on chromosome $4M^g$ of *Ae. geniculata* (Kynast et al., 2000). While the Gc action in each of these species results in the preferential transmission of each of these chromosomes there is evidence that the elements (alleles)/mechanisms responsible for the phenomenon between and within species are different (Endo, 1982; Endo, 1990).

The mode of action of gametocidal chromosomes is such that plants homozygous for the Gc chromosomes/elements do not show gametic

abortion. However, individuals that are heterozygous for the Gc element produce two types of gametes, i.e. gametes that either carry or lack the Gc element. Both male and female gametes that lack the Gc chromosome/elements undergo severe chromosome breakage (Finch et al., 1984; Nasuda et al., 1998) and, depending on the Gc chromosome/element, are generally not viable. As a result, only gametes carrying the Gc element are transmitted to the next generation. Thus, those traits linked with gametocidal elements, are found in all offspring without the requirement for selection, a genetic phenomenon that may have significant practical application in plant breeding (Endo and Gill, 1996; King et al., 1991).

The preferential transmission of the gametocidal chromosome $4S^{sh}$ in *Ae. sharonensis* has been hypothesized to be associated with two phenotypes controlled by at least two elements, i.e. the 'breaker' element and the 'inhibitor' element (Endo, 1990; Tsujimoto, 2005). The 'breaker' element, *Gc2* (Friebe et al., 2003) or *GcB* (Knight et al., 2015), causes extensive chromosomal breakage and hence lethality of gametes, which are generated from plants carrying a gametocidal chromosome, but do not themselves carry the gametocidal chromosome (Finch et al., 1984; King and Laurie, 1993; Nasuda et al., 1998). The role of the second element, the 'inhibitor' element, is to prevent the chromosomal breakage action of the 'breaker' element in gametes that have retained the Gc elements (Endo, 1990; Friebe et al., 2003; Tsujimoto, 2005).

Ae. sharonensis chromosome $4S^{sh}$ is homoeologous with group 4 in wheat and compensates well for regions of wheat chromosomes 4B (Friebe et al., 2003) and 4D (King et al., 1991) as well as whole chromosome substitutions for 4B and 4D in wheat (Miller et al., 1982). Friebe et al. (2003) developed a knock-out mutant of the 'breaker' element *Gc*2 on *Ae. sharonensis* chromosome $4S^{sh}$ (T4B-$4S^{sh}$#1, which has lost the chromosome 'breaker' function, but has retained the 'inhibitor' element). The *Ae. sharonensis* $4S^{sh}$ 'breaker' element has been previously mapped by C-banding to the distal end of the long arm (Endo, 2007) and has been shown, through deletion studies, to be limited to a small region immediately proximal to a sub-telomeric heterochromatin block at the distal end of the chromosome (Knight et al., 2015).

King et al. (1996) developed two hexaploid wheat introgression lines (Brigand 8/2 and Brigand 8/9), that involved the translocation of the distal end of the long arm of chromosome $4S^{sh}$ ($4S^{shL}$) from *Ae. sharonensis* onto the distal end of the long arm of chromosome 4D of *Triticum aestivum* (4DS·4DL-$4S^{shL}$ translocation). A strategy for mapping the Gc region of 4DS·4DL-$4S^{shL}$ translocation involving the irradiation of populations of these introgression lines and screening those with SNP markers specifically designed to the $4S^{shL}$ introgression segment for deletions has been undertaken (Knight et al., 2015). This work resulted in the development of 3 SNP markers within the Gc region that contained the *GcB* 'breaker' element.

The use of comparative mapping based on synteny, or the identical order of genetic loci between genomes of distantly related species, is an important approach for identifying gene positions and marker linkages in large genome species through comparisons with small genome model species like rice (Bennetzen and Freeling, 1993; Han et al., 1998; Moore et al., 1993). However, there are several potential limitations involved in this approach, including the observation that many traits of economic interest like disease resistance, and in our study the gametocidal elements, may be species specific and therefore not detectable or even absent in the simpler genomes. In addition, marker presence or order may not be conserved (Foote et al., 1997; Han et al., 1998) and marker polymorphism is often limited.

The genome of hexaploid wheat has been examined by targeted-capture sequencing, which is used to enrich for sequences of interest before carrying out NGS, using SureSelect, an in-solution based technology (Saintenac et al., 2011) and using a NimbleGen array (Winfield et al., 2012).

The aim of the current study was to further map the sub-telomeric region of $4S^{sh}$ consisting of the Gc elements, located on the 4DS·4DL-$4S^{shL}$ translocation, through characterisation of deletion mutants and perform comparative studies of that region of $4S^{shL}$ with wheat, rice and *Brachypodium distachyon*. Using the smaller introgression segment from line T4B-$4S^{sh}$#1 (Friebe et al., 2003) and comparative genomics, the Gc region was further characterised with 15 markers. We have also used

target-capture sequencing based on synteny between *Ae. sharonensis* and wheat to enable sequence investigation and variant calling in the $4S^{shL}$ region of the T4B-$4S^{sh}$#1 translocation segment carrying the knock-out mutation of *Gc2*, thereby providing 18 candidate gene sequences in *Ae. sharonensis* for the gametocidal 'breaker' element.

Materials and Methods

Plant material

Two *T. aestivum* cv. 'Brigand' introgression lines (ILs Brigand 8/2 and Brigand 8/9), each carrying a 4DS·4DL-$4S^{shL}$ translocation from *Ae. sharonensis* (AS) accession 2170001 (King et al., 1991), were individually crossed with *T. aestivum* cv. 'Huntsman', using Huntsman as the pollen donor, resulting in the generation of F_1 seeds that were hemizygous for the *Ae. sharonensis* introgression (Knight et al., 2015). Two deletion mutant populations were created by gamma irradiation of F_1 seed, with a dose 300 Gy, at the International Atomic Energy Agency's Laboratories (Seibersdorf, Austria; protocol no. FN662).

Other lines used in various experiments included *T. aestivum* cv. 'Chinese Spring' (CS); CS(4B)$4S^{sh}$ substitution line where wheat chromosomes 4B are missing and replaced with the *Ae. sharonensis* $4S^{sh}$ chromosomes (Miller et al., 1982); T4B-$4S^{sh}$#1 (Friebe et al., 2003), carrying a EMS mutation of the "breaker" element *Gc2*) and T4B-$4S^{sh}$ (Nasuda et al., 1998), the same translocation as T4B-$4S^{sh}$#1 but without the EMS

mutation). Line T4B-4S^{sh}#1 was kindly provided by Prof. Friebe, Department of Plant Pathology, Kansas State University, Manhattan, Kansas, USA. And line T4B-4S^{sh} was kindly provided by Dr. Nasuda, Laboratory of Plant Genetics, Kyoto University, Kyoto, Japan.

Fluorescence *in situ* hybridisation (FISH) and Genomic *in situ* hybridisation (GISH)

Preparation of chromosome spreads and the protocol for GISH was as described in King et al. (2017). For FISH, two repetitive DNA sequences pSc119.2 (McIntyre et al., 1990), and pAs.1 (Rayburn and Gill, 1986) were labelled with Alexa Fluor 488-5-dUTP (green coloration) and Alexa Fluor 594-5-dUTP (red coloration), respectively, and hybridized to the same set of slides sequentially.

Slides were initially probed using labelled genomic DNA of *Ae. sharonensis* (100ng) and fragmented genomic DNA of Chinese Spring (2000ng) as blocker to detect the *Ae. sharonensis* introgressions. Probe to block was in a ratio of 1 to 20 (the hybridization solution was made up to 10µl with 2 x SSC in 1 x TE). The slides were then bleached for multi-colour GISH and re-probed with labelled DNAs of *T. urartu* (100ng) and *Ae. taushii* (200ng) and fragmented DNA of *Ae. speltoides* (5000ng) as blocker in the ratio 1:2:50 to detect the AABBDD genomes of wheat.

Slide denaturation was carried out at 80 ^{o}C for 5 mins and the probes were hybridized at 55 ^{o}C for 16 h. Post hybridization, all slides were washed with 2 x SSC buffer for 1 min and counter-stained with DAPI. Slides were

examined using a Leica DM5500B epifluorescence microscope with filters for DAPI, Alexa Fluor 488 (FITC) and 594 (TRITC) and photographs were taken using a Leica DFC 350FX digital camera. For each metaphase, 3 exposures were taken, one each for DAPI, FITC and TRITC, and treated in order to decrease background noise and to increase the resolution of the signals. These images were combined into one final image using the Leica Application Suite package.

Marker development

One set of primer pairs were designed from wheat or barley ESTs (available in public databases) that had the best BLAST hit against coding sequences from the distal end of rice chromosome Os3S, *B. distachyon* chromosome Bd1L and also from *Ae. sharonensis* accession 1644 where coding sequences had been obtained from a transcriptome assembly by Bouyioukos et al. (2013). The sequence of the wheat BLAST hit result was BLASTN searched back against the rice or *B. distachyon* genome to confirm that the products were orthologous to the original sequence used for the BLAST search.

Another set of primer pairs were designed from the survey sequence of genes on the distal end of wheat 4BL/4DL (Mayer et al., 2014). Both sets of primer pairs were tested for polymorphisms between Brigand, Chinese Spring and *Ae. sharonensis* alleles on IL T4B-4S^{sh}#1 and IL 8/2 and 8/9.

All primers were initially tested by PCR amplification (using a touchdown program: 95 °C for 5 min, then ten cycles of 95 °C for 1 min, 65 °C for 30

s [−1 °C per cycle] and 72 °C for 2 min, followed by 40 cycles of 95 °C for 1 min, 58 °C for 30 s and 72 °C for 2 min) on genomic DNA and the amplification products were run on a 1.5% agarose gel. The amplified bands were cut from the gel, cleaned using the NucleoSpin™ Gel Extraction Kit (MACHEREY-NAGEL) and sequenced (Source Biosciences, UK) for SNP discovery. The resulting sequences were characterised for SNP markers between Brigand and *Ae. sharonensis* and these SNPs were confirmed by their presence in the sequences of IL T4B-4S^{sh}#1 and ILs 8/2 and 8/9. When SNP markers were confirmed they were converted, where possible, into high-throughput KASP™ assays (LGC Genomic, UK).

Marker names indicate the species source from which they were obtained, i.e. OsB8L10 (*Oryza sativa*), BD106 (*B. distachyon*), AS_17314 (*Ae. sharonensis*) and WHEAS1 (wheat; *T. aestivum*).

Genotyping and Phenotyping of Y300 M1 population

DNA was extracted from leaf material of M_1 plants at the 2–3 leaf stage, according to the Somers and Chao protocol http://maswheat.ucdavis.edu/PDF/DNA0003.pdf, original reference in (Pallotta et al., 2003).

All KASP™ assay amplification and analyses of the M_1 DNA samples was performed by LGC, UK and the results were viewed on SNPViewer software (LGC Group, UK). Genotyping through sequencing was performed on M_1 DNA samples with developed sequencing markers through PCR

amplification and subsequent sequencing of the amplification products as described above.

The observation of fertility of spikes confirmed the presence or absence of the gametocidal locus, i.e. lines retaining the wild type locus would be semi-sterile while the presence of fertile spikes indicated the potential loss of the gametocidal function (presumably as a result of a deletion induced by irradiation). Spike fertility was measured as the percentage of florets that contained well-developed seeds. Only the two outermost (oldest) florets within each spikelet were counted for a maximum of 4 spikes per M_1 plant. Undersized florets at the top and bottom of the spike were not included in the calculations. Parental plants of Huntsman and Brigand ILs 8/2 and 8/9 were grown as controls alongside each batch so that environmental effects on spike fertility could be taken into account. Adjustments were made to the fertility results by multiplying the mutant line fertility percentage by 100 and then dividing it by the fertility percentage for the controls. An average of approximately 50% viable seeds indicated that the 'breaker' element was present since it would cause abortion of around half of the gametes.

Comparative Mapping

Where *Ae. sharonensis* sequences were generated from mapped SNP or InDel markers or were obtained through the transcriptome assembly (Bouyioukos et al., 2013), the sequences were BLASTed against the

complete pseudomolecule (PM) assemblies for rice (Rice Genome Annotation Project v7.0; http://rice.plantbiology.msu.edu), *B. distachyon* (Phytozome v2.1; http://phytozome.jgi.doe.gov), the ordered PMs for wheat (IWGSC CSS v1.0; http://plants.ensembl.org/Triticum_aestivum) and against the barley genome assembly (IBSC v1.0; http://plants.ensembl.org/Hordeum vulgare) in order to generate comparisons with these species. In each case, only the most significant BLAST hit result (at least 90% sequence identity on at least 100 bp) was used to assign putative orthology between the *Ae. sharonensis* marker sequences and the PM sequences for the four comparison species. Figure 3 was drawn using Circos (Krzywinski et al., 2009) to compare the synteny between the virtual order of the *Ae. sharonensis* markers with orthologous genes identified in rice, *B. distachyon*, wheat and barley.

Capture assay design

Targeted sequence capture in hexaploid wheat Chinese Spring and in $4S^{shL}$ of *Ae. sharonensis* was performed using Agilent's SureSelect solution phase hybridization assay. The capture array was designed in collaboration with Agilent Technologies (Santa Clara, USA) and comprised 120-mer biotinylated oligonucleotide baits, each overlapping by 20 bases and thus achieving 2-fold coverage of the target sequences.

RNA baits were designed from a subset of 143 coding sequences from *Ae. sharonensis* transcriptome assembly sequences (Bouyioukos et al., 2013) homologous to genes at the distal end of wheat 4BL/4DL (Mayer et al.,

2014) for the Agilent SureSelect CustomXT kit. All wheat genomic DNA sequences were compared with each other to select only one representative homoeologous copy for each gene.

Construction of genomic DNA libraries, hybridization and sequencing

Four lines were included in the targeted sequence capture experiment: Chinese Spring; CS(4B)4S^{sh}; T4B-4S^{sh}#1 and T4B-4S^{sh}. Genomic DNA isolated from 3-week old seedlings (as described above) was used for library construction. DNA concentration was determined spectrophotometrically using a Nanodrop-1000 (Thermo Scientific, Pittsburgh, PA, USA).

DNA-library construction was performed by Source Biosciences, UK. For each genotype, 200 ng of genomic DNA dissolved in 50 µl of water was fragmented to an average size of 200 bp by 15 minutes of sonication on ice at maximum intensity (Virsonic 50, Virtis, Warminster, PA, USA). The following steps, i.e. fragment end-repairing, A-tailed ligation, adapter's ligation and final PCR, were performed according to the standard protocol of Agilent. The DNA fragments that had adapter molecules on both ends underwent 10 cycles of PCR to amplify the amount of prepared material. Amplified libraries were then individually hybridised and captured using the developed SureSelect Capture kit using Agilent's standard protocol. Captured libraries underwent 12 cycles of PCR to amplify the amount of captured DNA, and to also add the specific index sequences to each library.

The resulting libraries were then validated using the Agilent 2100 Bioanalyzer to confirm the molarity and size distribution. The libraries were diluted to 2nM and pooled in to one. The pool was validated using the Agilent 2100 Bioanalyzer to confirm the size distribution prior to sequencing. The average size of the libraries was 350bp. The pool was loaded at a concentration of 8pM onto an Illumina MiSeq flow cell v2. Samples were then sequenced using 150bp Paired-End runs on a MiSeq instrument. The unprocessed reads were submitted to the European Nucleotide Archive (http://www.ebi.ac.uk/ena) under project accession number PRJEB14397.

Sequence data trimming, filtering, assembly, variant calling and viewing alignments

The following work was performed by Source Biosciences, UK. Adaptor and quality trimming and filtering of Illumina reads was performed using Skewer (v.1). Trimming settings were adjusted for variant calling. The trimmed reads for T4B-4S^{sh} were assembled into contigs using CLC assembly cell (v.4.2.0) and used as a reference genome. The trimmed reads for the other samples were mapped onto this de-novo assembled reference sequence with Bowtie (v.2.2.5). In order to remove variant prediction bias caused by PCR artefacts, reads were de-duplicated by using Picardtools (v.1.130) prior to variant calling. De-duplicated reads in BAM format were then re-aligned around InDels to minimise misalignments using the Genome Analyis ToolKit (GATK) standard pipeline (v3.5). These

reads were then used for calling single nucleotide variants and small insertions or deletions present in each sample, when compared to the reference sequence, using the GATK Haplotype Caller. SNP calls were finally filtered using a minimum quality of 30 and depth of 2. The GATK's GenotypeGVCFs tool was then used to correlate genotype calls across the samples. Integrative Genomics Viewer (IGV v.2.3.58) was used to view the BAM alignment files to confirm the presence/absence of SNPs in samples mapped onto the reference genome.

Results

In situ hybridisation of the 4S^{shL} introgression

Two sets of *in situ* hybridisation experiments were performed to confirm the presence of *Ae. sharonensis* introgressions in the various ILs used in this study. Initially, the introgressions in ILs Brigand 8/2 and 8/9 were confirmed with treatment by FISH probes followed by multi-colour GISH. In the second experiment, the presence of the *Ae. sharonensis* introgression in IL T4B-4S^{sh} and its EMS-mutant counterpart was detected by a genomic DNA probe of *Ae. sharonensis* followed by multi-colour GISH to indicate the wheat chromosome containing the translocation.

The presence of the *Ae. sharonensis* introgression in *T. aestivum* cv. Brigand was confirmed by FISH and GISH of chromosome spreads at metaphase from root-tips as shown in Figure 1. *In situ* hybridisation was performed on Brigand wheat; ILs Brigand 8/2 and 8/9; a 4B-4S^{sh}

substitution line in Chinese Spring background CS(4B)4S^{sh} and on the *Ae. sharonensis* accession used for making the introgression and substitution lines. FISH probes pSc119.2, a rye repetitive DNA sequence (McIntyre et al., 1990) and pAs.1, a repetitive DNA sequence from *Ae. tauschii* (Rayburn and Gill, 1986) were used to distinguish between *Ae. sharonensis* and wheat chromosomes by analysing the different fluorescence patterns in the parental lines, *Ae. sharonensis* (Figure 1a) and Brigand (Figure 1d), and in the substitution line CS(4B)4S^{sh} (Figure 1c). This karyotypic analysis (as compared to Chinese Spring - Figure 1a; (Zhang et al., 2013)) enabled the identification of the pair of 4D chromosomes in Brigand and the translocated 4S^{shL} chromosome segment (in the 4D/4S^{shL} translocation) in ILs Brigand 8/2 and 8/9 (identified by arrows in Figures 1d, 1e and 1f, respectively). Together, the two FISH probes allowed identification of all 21 homologous chromosomes pairs in both lines of interest, IL Brigand 8/2 and IL Brigand 8/9, and this which was fully corroborated by simultaneous GISH analysis (Figures 1g-i).

The 4S^{shL} chromosome segment in ILs Brigand 8/2 and 8/9 showed a strong signal for probe pSc119.2 at its distal end (Figure 1e and 1f). This matches the one present at the distal ends of the chromosomes from *Ae. sharonensis* (Figure 1b). The same strong signal was also present on chromosome 4S^{sh} in the CS(4B)4S^{sh} substitution line (identified by arrows in Figure 1c) but is absent in chromosome 4D of Brigand (identified by arrows in Figure 1d). Multi-colour GISH, which separates the chromosomes of A (green), B (blue-purple) and D (red) genomes of wheat, showed the

presence of the *Ae. sharonensis* B-genome related genomic content at the distal end of chromosome 4D of ILs Brigand 8/2 and 8/9 (Figure 1h and 1i), which was absent in wild-type Brigand wheat (Figure 1g).

A different approach was used to detect the *Ae. sharonensis* introgressions in IL T4B-4S^{sh}. GISH with a genomic DNA probe of *Ae. sharonensis* was performed on *Ae. sharonensis*, the CS(4B)4S^{sh} substitution line and ILs T4B-4S^{sh} and T4B-4S^{sh}#1 followed by multi-colour GISH on these lines. Figure 2 shows the *Ae. sharonensis* chromosomes giving a strong signal when probed with its own genomic DNA (Figure 2a). The probe also clearly detects the 4S^{sh} chromosomes in the CS(4B)4S^{sh} substitution line as indicated by arrows in Figure 2b. Both ILs T4B-4S^{sh} and T4B-4S^{sh}#1 also showed the presence of *Ae. sharonensis*-derived chromatin, reported to be 4S^{ShL} (Friebe et al., 2003; Nasuda et al., 1998), at the distal ends of two wheat chromosomes as indicated by arrows in Figures 2c and 2d respectively. The 4S^{shL} translocations were confirmed to be at the distal ends of B-genome (blue-purple) chromosomes, reported to be chromosome 4B (Friebe et al., 2003; Nasuda et al., 1998), as indicated by multi-colour GISH in Figures 2e-f.

Characterisation of the gametocidal introgression from *Ae. sharonensis*

Six hundred and seventy new primer pairs were designed, 430 from wheat and barley ESTs (primer set 1) and 240 from the wheat survey sequence of chromosome 4BL/4DL (primer set 2), and tested for the production of

amplification products from all parental genotypes Brigand, Chinese Spring and *Ae. sharonensis* and the ILs Brigand 8/2, 8/9 and IL T4B-4S^{sh}#1. Approximately 10% of the primers designed were successful in producing a SNP (54) or an InDel (14) between the parental genotypes, which also transferred to all or some of the ILs. Of these 68 markers, only 15 markers (11 SNPs and 4 Indels) showed polymorphisms between the parent genotypes and all the ILs. The rest of the markers had polymorphisms that transferred to IL Brigand 8/2 and 8/9 but not to IL T4B-4S^{sh}#1. Out of the 54 SNP markers developed, 30 (21 from primer set 1 and 9 from primer set 2) were converted to KASP^{TM} assays and the remaining 24 SNPs (22 from primer set 1 and 2 from primer set 2) were used as sequencing markers (Table S1) to genotype the Y300 irradiated M_1 population.

A total of 2350 hybrid F_1 (M_0) seeds, from two crossing populations, were irradiated with a dose 300 Gy (1536 seeds from the IL Brigand 8/2 × Huntsman cross and 814 seeds from the IL Brigand 8/9 × Huntsman cross). The M_1 seeds germinated at a rate of 70.1% and these plants were further screened for deletions. From both Y300 mutant populations, 1658 plants were phenotyped for spike fertility and genotyped with 30 KASP^{TM} assays and 24 sequencing markers to reveal 16 lines carrying deletions of varying sizes in the 4S^{shL} segment as shown in Figure 3. The presence of a homozygous wheat allele (-/-) for the marker indicated that the marker was deleted in the 4S^{shL} introgression segment whereas the presence of a heterozygous call (+/-) showed that the marker was present within the introgression and hence the introgression segment at that region does not

have a deletion within it. The whole *Ae. sharonensis* segment had been deleted in two of these lines (lines 89_118 and 89_131, Figure 3) and thus they were omitted from further investigation. The *Ae. sharonensis* SNP markers were further ordered in the $4S^{shL}$ introgression segment by comparing the deleted markers in the different deletion segments and through comparative mapping with other crop species as subsequently explained.

Thirty of the 1658 M_1 plants showed high spike sterility where 2 or more spikes were completely sterile and 1614 lines showed semi-sterility (40-60% seed set). Of the 16 deletion lines identified, 4 showed restoration of fertility, i.e. 80-100% fertile spikes; 2 were sterile and the rest were semi-sterile (Figure 3). The parental ILs 8/2 and 8/9 showed fertile spikes due to homozygosity of the Gc elements and the T4B-$4S^{sh}$#1 line has been previously shown to have fertile spikes (Friebe et al., 2003).

The presence of only 11 SNP markers within the T4B-$4S^{sh}$#1 line (shown in yellow in Figure 3) indicated that the gametocidal loci-carrying introgression in this line is much smaller than the $4S^{shL}$ translocation in ILs 8/2 and 8/9 that carried all the SNP markers tested.

Synteny-based approaches to establish a putative gene order along *Ae. sharonensis* $4S^{shL}$

Using the information from the above genotyping on the M_1 population and the synteny between *Ae. sharonensis* $4S^{shL}$ marker sequences and genomic sequences of rice chromosome 3, *B. distachyon* chromosome 1, wheat

chromosome 4B and barley chromosome 4, we were able to propose a rough order of markers along the $4S^{shL}$ introgression segment. Among the 68 markers developed in the $4S^{shL}$ segment, 57 orthologous genes were identified in rice, 61 in *B. distachyon*, 38 in wheat and 51 in barley (Table S2). The 14 InDel markers that were not used for genotyping were ordered purely based on synteny with rice and *B. distachyon* since they gave the maximum orthologous gene hits.

Figures 4 and 5 show the comparison of collinearity between chromosome segments from all 4 species, individually with the *Ae. sharonensis* $4S^{shL}$ segment from ILs 8/2 and 8/9. Figure 4a shows the syntenic relationship between $4S^{shL}$ and rice Os3S with large ribbons showing significant synteny. However, twists within these ribbons indicate that the markers are reciprocally mapped and even though they retain collinearity within these ribbons, their non-consecutive placement along Os3S indicates potential gene re-arrangements in *Ae. sharonensis* $4S^{shL}$ compared to rice. Similarly Figure 4b shows significant synteny between *Ae. sharonensis* $4S^{shL}$ and *B. distachyon* Bd1L. However, groups of reciprocal markers (indicated by twisted ribbons) on $4S^{shL}$ are non-collinear compared to Bd1L as indicated by the various criss-cross patterns of ribbons but markers retain collinearity within the group. Eleven non-syntenic genes were found to be orthologous in rice and seven in *B. distachyon*.

Figures 5a and 5b show comparison in synteny between *Ae. sharonensis* $4S^{shL}$ and wheat (*T. aestivum*) chromosome 4B and barley chromosome 4 respectively. Small groups of markers (between 2-13) are syntenic

between these genomes as indicated by large ribbons and maintain collinearity. However, unlike rice and *B. distachyon*, there is potentially significant gene re-arrangement in *Ae. sharonensis* $4S^{shL}$, as compared to wheat and barley, indicated by individual coloured links, representing a single marker on $4S^{shL}$, that cross map to non-collinear positions on wheat and barley chromosomes. Two non-syntenic genes were found to be orthologous in wheat and 3 in barley.

Illumina sequencing and de novo assembly

Overlapping 120-mer RNA baits were designed to target approximately 1 Mb of sequence selected from *Ae. sharonensis* sequences (Bouyioukos et al., 2013) homologous to genes at the distal end of wheat 4BL/4DL (Mayer et al., 2014). Best blast hit analysis of genomic DNA of 153 genes at the distal end of wheat 4BL/4DL against the *Ae. sharonensis* whole-transcriptome assembly sequences (Bouyioukos et al., 2013) resulted in the identification of 143 *Ae. sharonensis* exonic sequences (Table S3). These sequences were used to design 11,924 unique probes (Table S4). A majority of the probes were replicated to produce the RNA bait library for the Agilent SureSelect CustomXT kit consisting of a total of 22,996 overlapping probes.

Table 1 shows the summary of sequencing and read-mapping data for each sample. The high-quality reads from T4B-$4S^{sh}$, spanning 26,181,499 bp were assembled into 66,517 contigs with an average length of 394 bp and an N_{50} length of 400 bp.

Variant Calling and Unique SNP discovery

In order to find SNPs unique to the EMS–mutated T4B-4S^{sh}#1 line, known to have a point mutation in the gametocidal 'breaker' element, variant calling was performed on this line using the non-mutated T4B-4S^{sh} line as a reference genome. 26,572 homozygous SNPs were discovered between the two lines of which 9,131 were of the EMS-type (G-A or C-T transitions; (Greene et al., 2003). When these SNP positions were compared to those obtained for CS(4B)4S^{sh} and Chinese Spring with the same reference genome, 1,231 were found to have the reference SNP allele in CS(4B)4S^{sh}, but were unmapped in Chinese Spring. These SNPs are therefore unique to T4B-4S^{sh}#1 and 523 (42.5%) were of EMS-type. Focusing on those SNPs that were located on chromosome 4BL or those in regions with no alignment to the Chinese Spring reference, only 22 candidate EMS SNP positions across 18 contigs were identified, resulting in 18 potential candidate genes in *Ae. sharonensis* for the 'breaker' element of the gametocidal locus. To distinguish EMS-induced mutations, that were present in one of the three lines, from homoeologous and species-specific SNPs, which would be present in all three lines, a filter was applied to identify mutations as those SNPs with at least 80% allele frequency in T4B-4S^{sh}#1, but no higher than 20% in the reference and CS(4B)4S^{sh} lines with a mapping depth of 0 in Chinese Spring and with at least one read supporting the variant call. The results of these tests are summarized in Table 2.

The SNPs detailed in Table 2 were defined as synonymous/non-synonymous where possible. The 18 reference contigs from which the SNPs were derived were used in a BLASTX alignment of the translated nucleotide query to the BLAST nr protein database (Altschul et al., 1990) to determine the potential for translation of the SNP region and the reading frame for translation as per previously reported methodologies (Gardiner et al., 2015). Six of the SNPs were predicted as non-synonymous using this methodology, their predicted protein changes are detailed in Table 2 and one such SNP is shown in Figure 6.

Discussion

The *in situ* hybridisation work indicates that the *Ae. sharonensis* $4S^{shL}$ genomic content is related to the B-genome. This is supported by the presence of the $4S^{shL}$ translocation on chromosome 4B of Chinese Spring in T4B-$4S^{sh}$ (Nasuda et al., 1998). However, recent work has indicated that some species in the S genome, including *Ae. sharonensis*, might be more closely related to the D genome of hexaploid wheat as compared to the B-genome as originally thought (Marcussen et al., 2014). This theory is supported by the presence of the $4S^{shL}$ translocation in IL Brigand 8/2 and 8/9 on chromosome 4D.

Previous work has been done to produce a genetic linkage map of *Ae. sharonensis* (Olivera et al., 2013). However, of the ten linkage groups produced using 389 DArT and SSR markers, four were assigned to chromosomes $1S^{sh}$, $2S^{sh}$, $3S^{sh}$ and $6S^{sh}$ but none were identified as chromosome $4S^{sh}$. Thus, it was difficult to establish a set of markers that could be used to characterise the $4S^{shL}$ introgression lines from the outset. Identification of the approximate gene content of the $4S^{shL}$ terminal region was attempted through synteny between *Triticeae* chromosome group 4, *B. distachyon* chromosome 1, and rice chromosome 3. Initially, the markers were ordered using this synteny. However, preliminary work done on mapping of SNP markers designed between *Ae. sharonensis* accessions 1644 and 2232 (resistant and susceptible to race TTSK (Ug99) of stem rust pathogen *Puccinia graminis* f. sp. *tritici*, respectively; (Bouyioukos et al., 2013)) produced a genetic map of $4S^{sh}$ (unpublished data) that was not completely collinear with rice, *B. distachyon* and barley. These significant differences in gene arrangements and deletions only gave an initial tentative marker order that was then further fine-tuned through mapping of the markers in the deletion mutants of the Y300 population. However, the initial marker order allowed selection of KASP™ assays that were spread evenly across the $4S^{shL}$ introgression segment.

To screen for novel deletion lines with different breakpoints and thereby define more closely the region containing the gametocidal elements, it was necessary to develop PCR markers for loci dispersed within the $4S^{shL}$ introgression. At the initiation of this project, little genomic sequence

information was available for wheat and therefore the primer set 1 design was based purely on cDNA sequences. Towards the latter half of the project, publication of the *Ae. sharonensis* transcriptome assembly and the wheat survey sequences allowed design of primer set 2. Out of the 68 markers developed, 54 were used to genotype the M_1 population in order to pick out deleted segments in the introgressions and order the markers further within the $4S^{shL}$ segment. After establishing a marker order that fitted well with synteny and the deletions, it was apparent that even though there is considerable conservation of synteny among the species, the micro-collinearity of genes within $4S^{shL}$ with rice, *B. distachyon*, wheat and barley was extensively rearranged within the terminal region of $4S^{shL}$, especially compared with the latter two genomes, due to inversions. This is consistent with the reported decrease in synteny levels in the distal regions of chromosomes (Akhunov et al., 2003). Conserved synteny enables the exploitation of plants with smaller genomes, such as rice and *B. distachyon*, as new sources of markers. The extensive comparative mapping used in this study has allowed the reconstruction of a map of $4S^{shL}$, containing the gametocidal loci.

The work showed that even though there was a considerable level of synteny at the micro level between *Ae. sharonensis* $4S^{shL}$ and wheat 4BL, considerable changes had occurred through inversions. Translocations involving the terminal regions of *Triticeae* chromosomes have been previously observed (Devos and Gale, 1992; Moore et al., 1993) and evidence indicates that the three ancestral genomes are themselves

diverging more rapidly in these regions (Devos and Gale, 1992). It would have been interesting to see how the gene collinearity compares between $4S^{shL}$ and the wheat 4DL chromosome to see if the S-genome of *Ae. sharonensis* is more closely related to the D-genome as shown by Marcussen et al. (2014) rather than the B-genome as was previously thought (Maestra and Naranjo, 1997). However very few BLAST hits were obtained on 4D that had ordered PM positions in order to make a comparative map.

The presence of a small introgression in T4B-$4S^{sh}$#1, known to have a point mutation in the gametocidal 'breaker' locus, as compared to that in ILs 8/2 and 8/9, allowed us to focus our analysis to a smaller region of the *Ae. sharonensis* genome. Spike fertility data of the deletion lines also supports the hypothesis that the Gc elements potentially exist in this smaller distal region of $4S^{shL}$. Lines 82_58, 89_9, 89_118 and 89_131, all with this region deleted, showed restoration of fertility indicating that the 'breaker' element, heterozygous presence of which causes semi-sterility, potentially exists in this distal region of $4S^{shL}$. Lines 82_16 and 82_148 had a partial deletion in this region of the chromosome, but were semi-sterile, which could help in further mapping of the Gc region of chromosome $4S^{shL}$. Lines 82_579 and 82_689 were two of thirty sterile lines in the M_1 population and also showed a deletion in the $4S^{shL}$ introgression. Sterility would be an indication of the potential deletion of the 'inhibitor' element however, irradiation could also have contributed to the sterile phenotype observed in these lines.

In order to find the mutation responsible for the fully fertile phenotype in line T4B-4S^{sh}#1, a comparison of the sequence of the mutant introgression was made with it's non-mutant sequence from the T4B-4S^{sh} line using target capture and next-generation sequencing. This work resulted in 18 potential candidate genes in *Ae. sharonensis* for the gametocidal 'breaker' element *GcB*.

However, none of the samples exhibited coverage over the full targeted region. Possible causes for this include errors in assembling the reference genome sequence, poor read mapping (for repeated regions, for example), or biased amplification or sequencing of genomic DNA. The target capture probes were designed based purely on coding sequence with no account for intron positions or intergenic content since these sequences are not ordered in *Ae. sharonensis*. The final set of 143 CDS sequences totalled approximately 1 Mbp but it was anticipated that the close sequence identity between homoeologous genes in wheat and orthologous genes between wheat and *Ae. sharonensis* would allow the capture of ~3 Mbp of gene space (assuming three homoeologues per gene). Approximately 26 Mb region was sequenced and assembled into contigs to form the reference genome. This is due to non-target sequence capture and assembly and could have contributed to poor map reading of the samples onto the reference genome. Since the target capture was designed from *Ae. sharonensis* coding sequence, it is possible the re-sequencing of wheat and wheat introgression lines with this exon capture can be inefficient due to sequencing of a high proportion of intron sequence relatively to exons. Even

though it has been suggested that intron size is not likely to be a major limiting factor in the success of exon capture in wheat (King et al., 2015), no such studies have been done in *Ae. sharonensis* to allow discounting large intron sequences as a contributing factor to the large amount of non-target sequence captured in the *Ae. sharonensis* segments in wheat. It has also been shown that GC content affects PCR amplification efficiency (Strien et al., 2013), and target regions with an overall GC content higher than 60% or lower than 30% exhibit a significantly lower coverage (Henry et al., 2014). In addition, optimal parameters for mutation detection can change on data quality, efficiency of capture, sample type and the specific goal of the experiment. We did anticipate that homologous and paralogous reads whose perfect target was absent from the reference might mis-align during read-mapping and SNP detection, generating false positives at a range of frequencies that might be difficult to distinguish from true heterozygous mutations. But since the lines tested are reported to be homozygous, we hoped this would maximise the proportion of homozygous alleles and thereby simplify the analysis of SNPs in exon capture data in this pilot-scale experiment.

EMS has been shown to lead to predominantly G-A and C-T transitions in the mutations detected across several wheat genes (Chen et al., 2012; Uauy et al., 2009). We were able to classify the resulting SNPs as EMS or non-EMS according to the variant base and found that 43.7% of high-confidence SNPs detected (those with read support 10 and allele frequency 80%), which were specific to T4B-4S^{sh}#1, were G-A or C-T transitions. This

value is inconsistent with previous reports that suggest that EMS is potentially more specific in wheat than in other species such as rice (Henry et al., 2014). However, this non-specificity can be attributed to the dilution of the detected SNP pool with heterozygous reads that are inherent within the three sub-genomes of wheat. Thus, this detection of homoelogous SNPs between reads of the reference sequence and T4B-4S^{sh}#1, enhanced by significant non-target capture sequences, could explain the relatively lower proportion of EMS-type SNPs in our pilot-scale results than expected.

The assessment of the effect of these 22 mutations on gene function yielded 6 SNPs that were predicted to change the translated protein sequence. One of these SNPs (contig_26579, position 67) showed sequence conservation with a pentatricopeptide repeat containing protein that is thought to be involved in transposition to unrelated chromosomal sites (Geddy and Brown, 2007). This finding might support the transposon theory suggested by Tsujimoto and Tsunewaki (Tsujimoto and Tsunewaki, 1985) which was based on hybrid dysgenesis in *Drosophila melanogaster*, caused by *P* elements, where the symptoms include sterility, lethality, mutations and chromosome breakage. Since these symptoms and the mechanism through which the mobile elements operate are very similar to gametocidal induced symptoms and the two-loci theory, Knight et al. (2015) discussed the hypothesis that the breaker may be a transposon similar to these telomeric *P* elements and the inhibitor would be located close to the breaker in the sub-telomeric region. Work is presently underway to confirm whether any

of the 18 candidate genes identified are responsible for the gametocidal action.

A further 23,968 homozygous SNPs were found between Chinese Spring and the reference genome. After applying previous parameters 1,618 homozygous SNPs were unique to Chinese Spring and hundreds of SNP markers were obtained between Chinese Spring and *Ae. sharonensis* within the $4S^{shL}$ region. These could be used in future deletion mapping experiments in case the above candidate genes were found to be off-target. Thus, targeted sequencing has provided us with a valuable source of sequence information and possible high-density SNPs to exhaust the distal end of $4S^{shL}$ allowing further investigation into the gametocidal locus. This approach could also be useful in future to generate abundant SNPs for characterisation and manipulation of introgressed segments in wheat from its wild-relatives.

Aegilops sharonensis has been reported to be a rich source of genetic diversity for biotic and abiotic stress tolerance. Studies have identified *Ae. sharonensis* accessions with resistance to many pathogens and pests, including powdery mildew, leaf rust, stem rust and stripe rust (Gill et al., 1985; Olivera et al., 2007; Valkoun et al., 1985) and Hessian fly and greenbug (Gill et al., 1985). *Ae. sharonensis* has also been identified to be salt tolerant (Xu et al., 1993). The single dominant gene (designated *LrAeSh1644*) conferring resistance to leaf rust race THBJ in *Ae. sharonensis* accession 1644 was reported to be on Chr $6S^{sh}$ (Olivera et al., 2013). However, there are no reports of any important genes, apart from the

gametocidal loci, on chromosome $4S^{sh}$. Numerous studies have successfully exploited the function of chromosomal breakage by Gc chromosomes to induce intergeneric translocations that import alien chromosome segments or useful genes of the wild relatives into wheat (Friebe et al., 2000; Kwiatek et al., 2016; Li et al., 2016; Liu et al., 2010; Luan et al., 2010; Masoudi-Nejad et al., 2002; Wang et al., 2003). The ability to induce chromosome exchange between non-collinear chromosomes will make a step change in widening the pool of wild species, which could be exploited in wheat breeding, a prospect deemed vital for the production of new, superior breeds of wheat.

In conclusion, we were able to putatively order 54 SNP markers on the *Ae. sharonensis* $4S^{shL}$ map, containing the gametocidal locus, using deletion mapping and comparative genomics and with hundreds more SNPs that can be used in future work to further fine map the Gc locus. To date we have identified 18 candidates for the gametocidal 'breaker' element.

Acknowledgements

This work was funded by the Biotechnology and Biological Sciences Research Council (BBSRC) under research grant BB/H012834/1. The funding body played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript. We are grateful to Prof. Bernd Friebe from Kansas State University, Dr. Shuhei Nasuda from Kyoto University and Steve Reader from John Innes Centre for providing seeds for lines of interest. We are also grateful to Dr. Matthew

Moscou from The Sainsburys laboratory and Dr. Burkhard Steuernagel and Dr. Brande Wulff from John Innes Centre for their scientific consultation. The authors declare that they have no conflict of interests.

References

Akhunov, E.D., Akhunova, A.R., Linkiewicz, A.M., Dubcovsky, J., Hummel, D., Lazo, G., Chao, S., Anderson, O.D., David, J., Qi, L., Echalier, B., Gill, B.S., Miftahudin, Gustafson, J.P., La Rota, M., Sorrells, M.E., Zhang, D., Nguyen, H.T., Kalavacharla, V., et al. (2003). Synteny perturbations between wheat homoeologous chromosomes caused by locus duplications and deletions correlate with recombination rates. Proc. Natl. Acad. Sci. USA 100:10836-10841. doi: 10.1073/pnas.1934431100

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. 215:403-410. doi: 10.1016/s0022-2836(05)80360-2

Bennetzen, J.L., and Freeling, M. (1993). Grasses as a single genetic system: genome composition, collinearity and compatibility. Trends Genet. 9:259-261. doi: 10.1016/0168-9525(93)90001-X

Bouyioukos, C., Moscou, M.J., Champouret, N., Hernandez-Pinzon, I., Ward, E.R., and Wulff, B.B. (2013). Characterisation and analysis of the *Aegilops sharonensis* transcriptome, a wild relative of wheat in the Sitopsis section. PLoS One 8:e72782. doi: 10.1371/journal.pone.0072782

Chen, L., Huang, L., Min, D., Phillips, A., Wang, S., Madgwick, P.J., Parry, M.A., and Hu, Y.G. (2012). Development and characterization of a new TILLING population of common bread wheat (*Triticum aestivum* L.). PLoS One 7:e41570. doi: 10.1371/journal.pone.0041570

Devos, K.M., and Gale, M.D. (1992). The use of random amplified polymorphic DNA markers in wheat. Theor. Appl. Genet. 84:567-572. doi: 10.1007/bf00224153

Endo, T.R. (1978). On the *Aegilops* chromosomes having gametocidal action on common wheat. Proceedings of the 5th International Wheat Genetics Symposium 1:306-314.

Endo, T.R. (1979). Selective gametocidal action of a chromosome of *Aegilops cylindrica* [a wild wheat] in a cultivar of common wheat. Wheat Inf. Serv. 50:24-28.

Endo, T.R. (1982). Gametocidal chromosomes of three *Aegilops s*pecies in common wheat. Can. J. of Genet. Cytol. 24:201-206. doi: 10.1139/g82-020

Endo, T.R. (1985). Two types of gametocidal chromosome of *Aegilops sharonensis* and *Ae. longissima*. Jpn. J. Genet. 60:125-135. doi: 10.1266/jjg.60.125

Endo, T.R. (1988). Induction of chromosomal structural changes by a chromosome of *Aegilops cylindrica* L. in common Wheat. J. Hered. 79:366-370.

Endo, T.R. (1990). Gametocidal chromosomes and their induction of chromosome mutations in wheat. Jpn. J. Genet. 65:135-152. doi: 10.1266/jjg.65.135

Endo, T.R. (2007). The gametocidal chromosome as a tool for chromosome manipulation in wheat. Chromosome Res. 15:67-75. doi: 10.1007/s10577-006-1100-3

Endo, T.R., and Gill, B.S. (1996). The Deletion Stocks of Common Wheat. J. Hered. 87: 295-307. doi: 10.1093/oxfordjournals.jhered.a023003

Endo, T.R., and Tsunewaki, K. (1975). Sterility of common wheat with *Aegilops triuncialis* cytoplasm. J. Hered. 66:13-18.

Endo, T.R., and Y., K. (1978). Finding of a selectively retained chromosome of *Aegilops caudata* L. in common wheat. Wheat Inf. Serv. 47-48:32-35.

Finch, R.A., Miller, T.E., and Bennett, M.D. (1984). "Cuckoo" *Aegilops* addition chromosome in wheat ensures its transmission by causing chromosome breaks in meiospores lacking it. Chromosoma 90:84-88. doi: 10.1007/bf00352282

Foote, T., Roberts, M., Kurata, N., Sasaki, T., and Moore, G. (1997). Detailed Comparative Mapping of Cereal Chromosome Regions Corresponding to the *Ph1* Locus in Wheat. Genetics 147:801-807.

Friebe, B., Kynast, R.G., and Gill, B.S. (2000). Gametocidal factor-induced structural rearrangements in rye chromosomes added to common wheat. Chromosome Res. 8:501-511.

Friebe, B., Zhang, P., Nasuda, S., and Gill, B.S. (2003). Characterization of a knock-out mutation at the *Gc2* locus in wheat. Chromosoma 111:509-517. doi: 10.1007/s00412-003-0234-8

Gardiner, L.J., Quinton-Tulloch, M., Olohan, L., Price, J., Hall, N., and Hall, A. (2015). A genome-wide survey of DNA methylation in hexaploid wheat. Genome Biol. 16:273. doi: 10.1186/s13059-015-0838-3

Geddy, R., and Brown, G.G. (2007). Genes encoding pentatricopeptide repeat (PPR) proteins are not conserved in location in plant genomes and may be subject to diversifying selection. BMC Genomics 8:130. doi: 10.1186/1471-2164-8-130

Gill, B.S., Sharma, H.C., Raupp, W.J., Browder, L.E., Hatchett, J.H., Harvey, T.L., Moseman, J.G., and Waines, J.G. (1985). Evaluation of *Aegilops* species for resistance to wheat powdery mildew, wheat leaf rust, hessian fly, and greenbug. Plant Disease 69:314-316.

Greene, E.A., Codomo, C.A., Taylor, N.E., Henikoff, J.G., Till, B.J., Reynolds, S.H., Enns, L.C., Burtner, C., Johnson, J.E., Odden, A.R., Comai, L., and Henikoff, S. (2003). Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in *Arabidopsis*. Genetics 164:731-740.

Han, F., Kleinhofs, A., Ullrich, S.E., Kilian, A., Yano, M., and Sasaki, T. (1998). Synteny with rice: analysis of barley malting quality QTLs

and rpg4 chromosome regions. Genome 41:373-380. doi: 10.1139/g98-027

Henry, I.M., Nagalakshmi, U., Lieberman, M.C., Ngo, K.J., Krasileva, K.V., Vasquez-Gross, H., Akhunova, A., Akhunov, E., Dubcovsky, J., Tai, T.H., and Comai, L. (2014). Efficient Genome-Wide Detection and Cataloging of EMS-Induced Mutations Using Exome Capture and Next-Generation Sequencing. Plant Cell 26:1382-1397. doi: 10.1105/tpc.113.121590

Kibirige-Sebunya, I., and Knott, D.R. (1983). Transfer of stem rust resistance to wheat from an *Agropyron* chromosome having a gametocidal effect. Can. J. of Genet. Cytol. 25:215-221. doi: 10.1139/g83-035

King, I.P., Cant, K.A., Law, C.N., Worland, A.J., Orford, S.E., Reader, S.M., and Miller, T.E. (1996). An assessment of the potential of 4DS.4DL-4s^{l}L translocation lines as a means of eliminating tall off types in semi-dwarf wheat varieties. Euphytica 89:103-106. doi: 10.1007/BF00015726

King, I.P., Koebner, R.M.D., Schlegel, R., Reader, S.M., Miller, T.E., and Law, C.N. (1991). Exploitation of a preferentially transmitted chromosome from *Aegilops sharonensis* for the elimination of segregation for height in semidwarf bread wheat varieties. Genome 34:944-949. doi: 10.1139/g91-146

King, I.P., and Laurie, D.A. (1993). Chromosome damage in early embryo and endosperm development in crosses involving the preferentially

transmitted 4S^{l} chromosome of *Aegilops sharonensis*. Heredity 70:52-59.

King, J., Grewal, S., Yang, C.Y., Hubbart, S., Scholefield, D., Ashling, S., Edwards, K.J., Allen, A.M., Burridge, A., Bloor, C., Davassi, A., da Silva, G.J., Chalmers, K., and King, I.P. (2017). A step change in the transfer of interspecific variation into wheat from *Ambylopyrum muticum*. Plant Biotechnol. J. 15:217-226. doi: 10.1111/pbi.12606

King, R., Bird, N., Ramirez-Gonzalez, R., Coghill, J.A., Patil, A., Hassani-Pak, K., Uauy, C., and Phillips, A.L. (2015). Mutation Scanning in Wheat by Exon Capture and Next-Generation Sequencing. PLoS ONE 10:e0137549. doi: 10.1371/journal.pone.0137549

Knight, E., Binnie, A., Draeger, T., Moscou, M., Rey, M.-D., Sucher, J., Mehra, S., King, I., and Moore, G. (2015). Mapping the 'breaker' element of the gametocidal locus proximal to a block of sub-telomeric heterochromatin on the long arm of chromosome 4Ssh of *Aegilops sharonensis*. Theor. Appl. Genet. 128:1049-1059. doi: 10.1007/s00122-015-2489-x

Krzywinski, M., Schein, J., Birol, İ., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: An information aesthetic for comparative genomics. Genome Res. 19:1639-1645. doi: 10.1101/gr.092759.109

Kwiatek, M., Majka, M., Ślusarkiewicz-Jarzina, A., Ponitka, A., Pudelska, H., Belter, J., and Wiśniewska, H. (2016). Transmission of the *Aegilops*

ovata chromosomes carrying gametocidal factors in hexaploid triticale (×Triticosecale Wittm.) hybrids. J. Appl. Genet. 57:305-315.

Kynast, R.G., Friebe, B., and Gill, B.S. (2000). Fate of multicentric and ring chromosomes induced by a new gametocidal factor located on chromosome 4M^{g} of *Aegilops geniculata*. Chromosome Res. 8:133-139.

Li, H., Lv, M., Song, L., Zhang, J., Gao, A., Li, L., and Liu, W. (2016). Production and Identification of Wheat-*Agropyron cristatum* 2P Translocation Lines. PLoS ONE 11:e0145928. doi: 10.1371/journal.pone.0145928

Liu, W.H., Luan, Y., Wang, J.C., Wang, X.G., Su, J.J., Zhang, J.P., Yang, X.M., Gao, A.N., and Li, L.H. (2010). Production and identification of wheat - *Agropyron cristatum* (1.4P) alien translocation lines. Genome 53:472-481. doi: 10.1139/g10-023

Luan, Y., Wang, X., Liu, W., Li, C., Zhang, J., Gao, A., Wang, Y., Yang, X., and Li, L. (2010). Production and identification of wheat-*Agropyron cristatum* 6P translocation lines. Planta 232:501-510. doi: 10.1007/s00425-010-1187-9

Maan, S.S. (1975). Exclusive preferential transmission of an alien chromosome in common wheat. Crop Sci. 15:287-292. doi: 10.2135/cropsci1975.0011183X001500030002x

Maestra, B., and Naranjo, T. (1997). Homoeologous relationships of *Triticum sharonense* chromosomes to *T. aestivum*. Theor. Appl. Genet. 94:657-663. doi: 10.1007/s001220050463

Marcussen, T., Sandve, S.R., Heier, L., Spannagl, M., Pfeifer, M., Jakobsen, K.S., Wulff, B.B.H., Steuernagel, B., Mayer, K.F.X., and Olsen, O.-A. (2014). Ancient hybridizations among the ancestral genomes of bread wheat. Science 345:1250092.

Masoudi-Nejad, A., Nasuda, S., McIntosh, R.A., Endo, T.R., and Endo, T.R. (2002). Transfer of rye chromosome segments to wheat by a gametocidal system. Chromosome Research 10:349-357. doi: 10.1023/a:1016845200960

Mayer, K.F.X., Rogers, J., Doležel, J., Pozniak, C., Eversole, K., Feuillet, C., Gill, B., Friebe, B., Lukaszewski, A.J., Sourdille, P., Endo, T.R., Kubaláková, M., Číhalíková, J., Dubská, Z., Vrána, J., Šperková, R., Šimková, H., Febrer, M., Clissold, L., et al. (2014). A chromosome-based draft sequence of the hexaploid bread wheat *Triticum aestivum* genome. Science 345:1251788. doi: 10.1126/science.1251788

McIntyre, C.L., Pereira, S., Moran, L.B., and Appels, R. (1990). New Secale cereale (rye) DNA derivatives for the detection of rye chromosome segments in wheat. Genome 33:635-640.

Miller, T.E., Hutchinson, J., and Chapman, V. (1982). Investigation of a preferentially transmitted *Aegilops sharonensis* chromosome in wheat. Theor. Appl. Genet. 61:27-33. doi: 10.1007/bf00261506

Moore, G., Gale, M.D., Kurata, N., and Flavell, R.B. (1993). Molecular Analysis of Small Grain Cereal Genomes: Current Status and Prospects. Nat. Biotech. 11:584-589.

Nasuda, S., Friebe, B., and Gill, B.S. (1998). Gametocidal genes induce chromosome breakage in the interphase prior to the first mitotic cell division of the male gametophyte in wheat. Genetics 149:1115-1124.

Olivera, P.D., Kilian, A., Wenzl, P., and Steffenson, B.J. (2013). Development of a genetic linkage map for Sharon goatgrass (Aegilops sharonensis) and mapping of a leaf rust resistance gene. Genome 56:367-376. doi: 10.1139/gen-2013-0065

Olivera, P.D., Kolmer, J.A., Anikster, Y., and Steffenson, B.J. (2007). Resistance of Sharon Goatgrass (Aegilops sharonensis) to Fungal Diseases of Wheat. Plant Disease 91:942-950. doi: 10.1094/PDIS-91-8-0942

Pallotta, M.A., Asayama, S., Reinheimer, J.M., Davies, P.A., Barr, A.R., Jefferies, S.P., Chalmers, K.J., Lewis, J., Collins, H.M., Roumeliotis, S., Logue, S.J., Coventry, S.J., Lance, R.C.M., Karakousis, A., Lim, P., Verbyla, A.P., and Eckermann, P.J. (2003). Mapping and QTL analysis of the barley population Amagi Nijo x WI2585. Aust. J. Agric. Res. 54:1141-1144.

Rayburn, A.L., and Gill, B.S. (1986). Molecular identification of the D-genome chromosomes of wheat. J. Hered. 77:253-255.

Saintenac, C., Jiang, D., and Akhunov, E.D. (2011). Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome. Genome Biol. 12:R88. doi: 10.1186/gb-2011-12-9-r88

Strien, J., Sanft, J., and Mall, G. (2013). Enhancement of PCR amplification of moderate GC-containing and highly GC-rich DNA sequences. Mol. Biotechnol. 54:1048-1054. doi: 10.1007/s12033-013-9660-x

Tsujimoto, H. (1994). Two new sources of gametocidal genes from *Aegilops longissima* and *Ae. sharonensis*. Wheat Inf. Serv. 79:42-46.

Tsujimoto, H. (2005). Gametocidal genes in wheat as the inducer of chromosome breakage. Frontiers of Wheat Bioscience Memorial Issue, Wheat Inf. Serv.:33-48.

Tsujimoto, H., and Tsunewaki, K. (1984). Gametocidal genes in wheat and its relatives. I. Genetic analyses in common wheat of a gametocidal gene derived from *Aegilops speltoides*. Can. J. of Genet. Cytol. 26:78-84. doi: 10.1139/g84-013

Tsujimoto, H., and Tsunewaki, K. (1985). Hybrid dysgenesis in common wheat caused by gametocidal genes. Jpn. J. Genet. 60:565-578. doi: 10.1266/jjg.60.565

Uauy, C., Paraiso, F., Colasuonno, P., Tran, R.K., Tsai, H., Berardi, S., Comai, L., and Dubcovsky, J. (2009). A modified TILLING approach to detect induced mutations in tetraploid and hexaploid wheat. BMC Plant Biol. 9:1-14. doi: 10.1186/1471-2229-9-115

Valkoun, J., Hammer, K., Kučerová, D., and Bartoš, P. (1985). Disease resistance in the genusAegilops L. — stem rust, leaf rust, stripe rust, and powdery mildew. Die Kulturpflanze 33:133-153. doi: 10.1007/bf01997267

Wang, X.P., Chu, J.H., and Zhang, X.Q. (2003). Efficient production of wheat alien translocation lines and characterization by molecular cytogenetics. Acta Genet. Sin. 30:619-624.

Winfield, M.O., Wilkinson, P.A., Allen, A.M., Barker, G.L., Coghill, J.A., Burridge, A., Hall, A., Brenchley, R.C., D'Amore, R., Hall, N., Bevan, M.W., Richmond, T., Gerhardt, D.J., Jeddeloh, J.A., and Edwards, K.J. (2012). Targeted re-sequencing of the allohexaploid wheat exome. Plant Biotechnol. J. 10:733-742. doi: 10.1111/j.1467-7652.2012.00713.x

Xu, X., Monneveux, P., Damania, A.B., and Zaharieva, M. (1993). Evaluation for salt tolerance in genetic resources of Triticum and Aegilops species. Plant Genet. Resour. Newsl. 96:11-16.

Zhang, H., Bian, Y., Gou, X., Zhu, B., Xu, C., Qi, B., Li, N., Rustgi, S., Zhou, H., Han, F., Jiang, J., von Wettstein, D., and Liu, B. (2013). Persistent whole-chromosome aneuploidy is generally associated with nascent allohexaploid wheat. Proc. Natl. Acad. Sci. USA 110:3447-3452. doi: 10.1073/pnas.1300153110

Figure captions

Figure 1. Root-tip metaphase chromosomes of wheat-*Ae. sharonensis* ILs Brigand 8/2 and 8/9 carrying disomic recombinant chromosomes.

(a) Chromosomes from Chinese Spring wheat, from A, B and D genomes, showing sites of *in situ* hybridisation with *pSc119.2* in *green* and *pAs.1* in *red*. The DAPI staining showing chromosome morphology is displayed in *blue* to improve contrast. (b) FISH image of *Ae. sharonensis* chromosomes (c) FISH image of CS(4B)4S^{sh} substitution line showing presence of the chromosome 4S^{sh} as indicated by the arrows (d) FISH image of wild-type Brigand wheat where chromosome 4D is indicated by arrows (e) FISH image of IL 8/2 showing the 4D/4S^{shL} introgression indicated by arrows (f) FISH image of IL 8/9 showing the 4D/4S^{shL} introgression indicated by arrows (g) GISH image to show wild-type chromosomes of A (green), B (blue) and D (red) genomes in Brigand wheat (h) GISH image to show the 4D/4S^{shL} introgression in IL 8/2 as indicated by arrows (i) GISH image to show the 4D/4S^{shL} introgression in IL 8/9 as indicated by arrows. Fluorescence-labelled genomic DNA from *T. urartu* (A-genome donor), *Ae. speltoides* (putative B-genome donor) and *Ae. tauschii* (D-genome donor) were used as probes for multi-colour GISH.

Figure 2. Root-tip metaphase chromosomes of wheat-*Ae. sharonensis* ILs T4B-4S^{sh} and T4B-4S^{sh}#1 carrying disomic recombinant chromosomes.
(a) Single-colour GISH image of chromosomes from *Ae. sharonensis* showing a strong hybridisation signal with its own genomic DNA probe (b) Single-colour GISH image of CS(4B)4S^{sh} substitution line showing presence of the chromosome 4S^{sh} as indicated by the arrows (c) Single-colour GISH image of IL T4B-4S^{sh} where a segment of 4S^{shL} is indicated by arrows (d)

Single-colour GISH image of IL T4B-4S^{sh}#1 showing the 4S^{shL} introgression indicated by arrows (e) Multi-colour GISH image to show chromosomes of A (green), B (blue) and D (red) genomes in IL T4B-4S^{sh} with the 4B/4S^{shL} introgression is indicated by arrows (f) Multi-colour GISH image to show the 4B/4S^{shL} introgression in IL T4B-4S^{sh}#1 as indicated by arrows. Fluorescence-labelled genomic DNA from *Ae. sharonensis* was used as probe for single-colour GISH and from *T. urartu* (A-genome donor), *Ae. speltoides* (putative B-genome donor) and *Ae. tauschii* (D-genome donor) were used as probes for multi-colour GISH.

Figure 3. Genotypes of ILs and M_1 deletion lines using 30 KASP^{TM} assays and 24 sequencing markers with respective spike fertility data.

‘-/-‘ in *blue* represents the homozygous wheat allele and ‘+/-‘ in *green* represents the *Ae. sharonensis* allele, heterozygous with the wheat allele. The markers in *yellow* at the bottom represent the smaller size of the 4S^{shL} introgression in IL T4B-4S^{sh}#1. Below each line is the respective percentage of viable seeds in spikes of these lines and is represented as percentage of spike fertility.

Figure 4. Schematic representation of the syntenic relationships between *Ae. sharonensis* 4S^{shL} segment and the orthologous chromosomes in rice and *B. distachyon*.

(a) rice Os3S, (b) *B. distachyon* Bd1L chromosomes in which syntenic links were identified with the 4S^{shL} segment are represented with orange and

green coloured lines respectively. Ribbons represent the syntenic regions identified with each species while twists within the ribbons indicate reciprocally mapped groups of markers. Markers names are indicated as ordered on the $4S^{shL}$ segment and corresponding orthologous gene names for both species is indicated where available. Black regions of the chromosomes correspond to the telomeres of *Ae. sharonensis* $4S^{shL}$, rice Os3S and *B. distachyon* Bd1L. The yellow segment around the distal end of $4S^{shL}$ represents the smaller introgression region in IL T4B-$4S^{sh}$#1.

Figure 5. Schematic representation of the syntenic relationships between *Ae. sharonensis* $4S^{shL}$ segment and the orthologous chromosomes in wheat and barley.

(a) wheat 4BL, (b) barley 4HL chromosomes in which syntenic links were identified with the $4S^{shL}$ segment are represented with blue and lilac coloured lines respectively. Large ribbons and individual coloured lines represent the syntenic regions identified with each species while twists within the ribbons indicate reciprocally mapped groups of markers. Markers names are indicated as ordered on the $4S^{shL}$ segment and corresponding orthologous gene names for both species is indicated where available. Black regions of the chromosomes correspond to the telomeres of *Ae. sharonensis* $4S^{shL}$, wheat 4BL and barley 4HL. The yellow segment around the distal end of $4S^{shL}$ represents the smaller introgression region in IL T4B-$4S^{sh}$#1.

Figure. 6 Visualisation and comparison of sequence of $4S^{shL}$ to find SNPs specific to EMS-mutant line T4B-$4S^{sh}$#1.

Snapshot of a region of sequence comparison at contig_64128 position 139, between reference sequence T4B-$4S^{sh}$ and Chr 4B substitution line CS(4B)$4S^{sh}$, both showing base G at that position as compared to the EMS-mutant line T4B-$4S^{sh}$#1 showing base A at that position. Since the sequence for Chinese spring does not exist for this contig, the SNP at position 139 is unique to T4B-$4S^{sh}$#1.

Supplemental Information

Table S1

Supplementary data giving primer details of 30 KASP assays and 24 sequencing markers developed for deletion mapping.

Table S2

Supplementary data showing BLAST results for all 68 polymorphic markers developed against *Brachypodium distachyon*, *Oryza sativa*, *Triticum aestivum* and *Hordeum vulgare*.

Table S3

Sequence details of 143 *Ae. sharonensis* contigs used to design the target-capture sequencing experiment.

Table S4

Sequence details of 11,924 RNA probes designed to produce the bait library of the Agilent SureSelect Custom XT kit.

Tables

Table 1. Summary of sequencing data from Chinese Spring (CS); CS(4B)4S^{sh}; T4B-4S^{sh}#1 and T4B-4S^{sh}.

	T4B-4S^{sh}	T4B-4S^{sh}#1	CS(4B)4S^{sh}	Chinese Spring
Total reads after trimming	17,765,669	9,812,102	11,867,736	7,246,868
Q30[a] (%)	86.29	92.86	93.32	92.82
Mapped reads to reference (%)	N/A	50.88	47.64	53.74

[a] indicates the percentage of sequences at a sequencing error rate of less than 0.1%

Table 2. Summary of variant calling details for 18 candidate contigs that showed unique SNPs between T4B-4S^{sh}#1 and T4B-4S^{sh}

Contig name	Position	REF	ALT	Raw depth read	Variant allele frequency (%)	SNP region	Predicted Protein change	Associated protein (BLASTX)
contig_14317	409	G	A	4	100.0	Exon	-	Hypothetical protein TRIUR3_21892 [*T.urartu*]
contig_18374	248	G	A	13	100.0	Intron	-	-
contig_26579	67	C	T	2	100.0	Exon	Proline→ Serine	Pentatricopeptide repeat-containing protein (PPR) At2g15630-like [*B.distachyon*]
contig_27312	157	C	T	3	100.0	Exon	Proline→ Serine	Wall-associated receptor kinase 3 [*Ae.tauschii*]
contig_36992	207	C	T	5	100.0	Exon	-	Hypothetical protein F775 19782 [*Ae.tauschii*]
contig_38090	148	C	T	12	100.0	Exon	Valine→ Isoleucine	Hypothetical protein F775 21897 [*Ae.tauschii*]
contig_39789	78	G	A	3	100.0	Intron	-	-
contig_39789	84	G	A	4	100.0	Intron	-	-
contig_39798	222	G	A	7	100.0	Intron	-	-
contig_40907	117	G	A	3	100.0	Exon	Alanine→ Threonine	Lectin-domain containing receptor kinase A4.2 [*Ae.tauschii*]
contig_45712	328	C	T	4	100.0	Intron	-	-
contig_45712	354	G	A	3	100.0	Intron	-	-
contig_49742	214	G	A	3	100.0	Intron	-	-
contig_56325	66	G	A	6	100.0	Intron	-	-
contig_57351	254	G	A	3	100.0	Intron	-	-
contig_58421	172	C	T	6	100.0	Exon	Glutamic acid→Lysine	GDSL esterase/lipase At5g45910-like Si006718m [*S. italica*]
contig_62039	270	G	A	3	100.0	Intron	-	-
contig_62723	298	C	T	5	100.0	Intron	-	-
contig_62723	337	C	T	6	100.0	Intron	-	-
contig_62723	357	G	A	5	100.0	Intron	-	-
contig_64128	139	G	A	12	80.0	Exon	Cysteine→ Tyrosine	Hypothetical protein F775 00528 [*Ae.tauschii*]
contig_64143	143	C	T	4	100.0	Intron	-	-